Exhibit IND14

Journal of Computer Assisted Tomography
17(2):200–205, March/April

Accuracy and Reproducibility in Volumetric Analysis of Multiple Sclerosis Lesions

Edward F. Jackson, Ponnada A. Narayana, Jerry S. Wolinsky, and Timothy J. Doyle

Abstract: The accuracy and reproducibility of dual-contrast segmentation based on nonparametric feature map analysis have been investigated in a multicomponent gelatin phantom. The root mean square errors in volume ranged from 0.02 cm^3 for small volumes to 3.8 cm^3 for larger volumes, with a mean error of 0.97 cm^3. Average inter- and intraobserver coefficients of variation were found to be <7% for all compartments. To evaluate the reproducibility of segmentation of clinical image data, volumes of total brain, CSF, and multiple sclerosis (MS) lesions were obtained from five image sets of MS patients. Inter- and intraobserver coefficients of variations were computed for the patient data and were found to be <5% for brain, 17% for CSF, and 20% for MS lesions. Such variations were found to be reduced by appropriate preprocessing of the images. **Index Terms:** Magnetic resonance imaging—Volumetric analysis—Multiple sclerosis—Brain, anatomy.

Although MRI has revolutionized the evaluation of demyelinating disease states such as multiple sclerosis (MS), quantitative volumetric analysis of lesion load is only now becoming feasible in a clinical setting. Previous volumetric efforts generally utilized techniques such as region-of-interest (ROI) computations, which are typically performed manually by a trained operator using a track ball or similar interface. Although several groups have applied this technique to analyze MR data in a number of disease states (1–8), these studies are generally extremely time consuming, require well-trained operators, and suffer from significant inter- and intraobserver variations in the computed volumes (2). Thresholding, histogram, and linear combination techniques, sometimes combined with preliminary ROI outlining, work quite well in cases where interfaces between tissues demonstrate distinct contrast changes and have also been applied to MR image analysis (9–14). However, these techniques are also time consuming, require significant operator interaction and training, and do not take full advantage of the range of contrast obtained with MRI.

Multispectral or feature map techniques, first proposed for MR image segmentation in 1985 (15), use complementary contrast images to define tissues. For MRI these images are typically acquired as dual-echo data sets that provide, for example, proton density-weighted (PDW) and T2-weighted (T2W) images simultaneously. Using these data sets, the operator defines tissues simply by sampling points within the tissues and a scatter plot of PDW pixel intensity vs. T2W pixel intensity allows for identification of clusters representing the various tissues (16,17). The tissue sampling session can be completed rapidly (within 10–20 min) after which the image segmentation proceeds without operator intervention, thereby reducing bias and, presumably, inter- and intraoperator variability. Although used primarily for three-dimensional imaging in applications such as surgical planning (16,18), these techniques have been applied to the study of brain changes in aging and Alzheimer's dementia (19) and preliminary results have been reported in the study of MS (20). Such a dual-contrast technique greatly aids in volumetric studies of MS. For example, on T2W images periventricular MS lesions may be difficult to separate from CSF, but CSF and brain parenchyma are clearly resolved. On PDW images, however, MS lesions are generally distinct from

From the Departments of Radiology (E. F. Jackson, P. A. Narayana, and T. J. Doyle) and Neurology (J. S. Wolinsky), University of Texas Medical School at Houston, Houston, TX. Address correspondence and reprint requests to Dr. E. F. Jackson at Department of Radiology, University of Texas Medical School, 6431 Fannin, 2.132 MSB, Houston, TX 77030, U.S.A.

CSF, but CSF and brain parenchyma may not be as clearly resolved as they are on the T2W images. The complementary images provide improved separation of the tissues of interest.

Quantitative determination of lesion burden provides a means for establishing a parametric index for the follow-up of MS patients and for comparison to clinical data such as neuropsychological evaluations (21). In addition, the ability to accurately follow the volumes of brain, CSF, and lesions in a longitudinal manner using semiautomatic segmentation should allow for a more systematic approach to the determination of the effect of therapy and natural progression of the disease state. However, few studies have been performed to verify the accuracy and reproducibility of the feature map segmentation algorithm for volumetric analysis (18,20). The purpose of this study, therefore, was to evaluate the accuracy and inter/intraobserver variability in the determination of volumes using the feature map technique for semiautomated segmentation of MS patient images.

MATERIALS AND METHODS

Image Acquisition

All MR images were obtained using a GE 1.5 T Signa imager (GE Medical Systems, Waukesha, WI, U.S.A.). Dual-echo data sets were acquired from 3 mm contiguous slices with one average using TEs of 20 ms and 70 ms and a TR of 2,500 ms. The acquisition matrix size was 256 × 192, the field of view was 24 cm, and the acquisition time was 17 min.

Image Analysis

The image data sets were segmented into the tissues of interest by four different observers using a modified version of the software described by Cline et al. (16,18) implemented on a SPARCstation 2 workstation (Sun Microsystems, Mountain View, CA, U.S.A.). The software modifications were primarily related to pixel summing and volume calculations for the tissues of interest. With respect to the segmentation algorithm, the level of experience of the observers ranged from very experienced to novice. Each observer chose representative pixels from the tissues of interest (>30 points per tissue) and the pixel intensities were plotted as a scatter plot in feature space. As there was no physical basis for assuming the feature space clusters would satisfy a particular distribution function, the feature map used to segment the tissues was constructed from the identified tissue points using the nonparametric nearest neighbor algorithm (22,23). The total volumes of the tissues of interest were computed by summing the segmented pixels and multiplying by the previously measured in-plane resolution and slice thickness. To allow for the calculation of intraobserver variations, each observer repeated the segmentation on four different occasions.

Phantom Studies

To test the accuracy of the segmentation technique, phantom studies were performed using a cylindrical phantom consisting of irregularly shaped gel pieces ("lesions") embedded in a matrix gel ("brain"). The matrix gel was a gelatin/ethylene glycol base doped with $MnCl_2$ and cross-linked by formaldehyde as described by Companion (24). The gelatin and $MnCl_2$ concentrations in the matrix gel were chosen such that the T1 and T2 relaxation times were 880 and 100 ms, respectively, to approximately simulate brain. The irregularly shaped lesion gels were formed from $MnCl_2$-doped gelatin but without cross-linking. For these pieces, the gelatin and $MnCl_2$ concentrations were chosen such that the T1 and T2 values were 1,600 and 220 ms, respectively, to approximately simulate MS lesion values. Three such lesion gels, of volumes 0.31, 0.21, and 0.10 cm^3, were each coated with a thin layer of paraffin to prevent diffusion and placed within the matrix gel. The irregular shapes of the lesions were specifically chosen to more realistically represent MS lesions as seen on MRI, and the small volumes were chosen to more closely model the problems associated with partial volume averaging. The phantom was scanned using the acquisition parameters above, and the resulting images segmented into "brain" and "lesion" plus background by the four different observers.

Patient Studies

To evaluate the inter- and intraobserver variations in segmentation of MS lesion, brain, and CSF volumes, five image data sets were acquired from patients with clinically definite MS (all women, age range 26–44 years). Axial images were acquired from the foramen magnum to the vertex of the brain using the parameters given above and were segmented into four tissues: brain (gray and white matter), CSF, lesions, and other (muscle, scalp, subcutaneous fat, etc.) plus background. For the training session, at least 30 points were sampled for each tissue from multiple axial planes; this task was accomplished in 10–20 min, dependent on experience. Following segmentation, the pixel counts from all slices for each tissue were summed and, using the same method described above, converted to volumes of the total brain, total CSF, and total lesions.

Each data set was analyzed by four different observers on four different occasions.

All patient studies were approved by the institutional Committee for Protection of Human Subjects.

Data Analysis

Intraobserver reproducibilities were expressed in terms of the coefficient of variation for the four volume determinations performed by each observer. For the interobserver reproducibility, the mean values of the computed volumes for each observer (averaged over the four repetitions) were first determined and these mean values were then used to compute the interobserver coefficient of variation. All coefficients of variation are expressed in terms of percent standard deviation of the mean. Univariate analysis of variance was performed using a standard statistical analysis package (Statgraphics; STSC, Rockville, MD, U.S.A.) to establish the significance of inter- and intraobserver variation.

RESULTS

Phantom Studies

Dual-echo images from a slice through the 0.21 cm^3 lesion gel of the phantom are shown in Fig. 1a (PDW) and b (T2W). The feature map obtained from the training points is shown in Fig. 1c where dark gray represents the portion of feature space corresponding to lesion gels and light gray represents the matrix gel. Figure 1d is the segmented image corresponding to this feature map and demonstrates the quality of segmentation obtained. The computed volumes of the lesion gels and matrix gel as well as the coefficients of variation and root mean square (RMS) percent error are given in Table 1 for each of the four observers. In general, the volumetric analysis underestimated the actual volumes slightly, but for each lesion gel the maximum RMS error was <0.05 cm^3 and the mean error was <0.03 cm^3. The mean RMS percent error for the matrix gel was 1.9% (3.8 cm^3). The average RMS error for all gel compartments was 0.97 cm^3. In terms of reproducibility, the intraobserver coefficient of variation for the lesion gels ranged from 2.5 to 15.2% with a mean value of 6.5%, while the interobserver coefficient of variation ranged only from 4.3 to 7.8% (mean 5.7%). The mean intra- and interobserver coefficients of variation for the matrix gel were 1.5 and 1.0%, respectively. Univariate analysis of variance indicated no significant interobserver variation in volume determination for any gel compartment ($p < 0.05$).

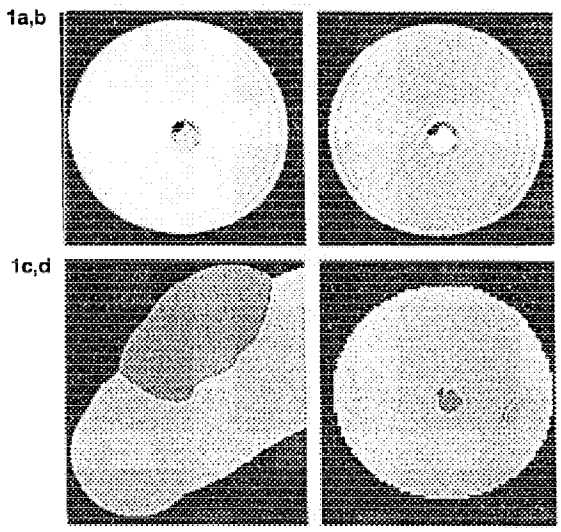

FIG. 1. One of the image planes from the phantom data set. **a:** Proton density-weighted image. **b:** T2-weighted image. **c:** Feature map. **d:** Segmented image. For the feature map and segmented image, light gray represents the matrix gel and dark gray represents the lesion gel.

Patient Studies

A typical segmented image taken from one of the patient image sets is given in Fig. 2. Figure 2a and b presents the PDW and T2W images, respectively, Fig. 2c is the feature map, and Fig. 2d is the segmented image. As can be seen from Fig. 2d, there are a few misclassified pixels, primarily in the scalp and bone marrow, which could be eliminated using connectivity algorithms (25). Due to the relatively few misclassified pixels, however, these algorithms were not initially applied. The pooled results of the volumetric analysis (all four repetitions from each of four observers) for the five image sets are given in Table 2. For these patients, the mean intraobserver coefficients of variation for brain, CSF, and MS lesions were 3.6, 15.5, and 20.3%, respectively, while the interobserver coefficients of variation were 4.2, 16.8, and 21.2%, respectively. Analysis of variance indicated a significant interobserver variation in brain volume determination for all image sets ($p < 0.05$). Significant interobserver variation was also found in CSF determinations in four of five data sets and in lesion volume determinations in two of five data sets. As might be expected, measurements from the more experienced operators demonstrated less interobserver variation. There was no significant intraobserver variation in any volume determination from any data set ($p < 0.05$).

DISCUSSION

For volumetric MRI to be used as a quantitative index of MS disease state, the accuracy and repro-

TABLE 1. *Computed volumes (V; cm³), coefficients of variation (CV; %), and RMS error (%) for four observers with four repetitions*

Observer no.	V_{matrix}	CV_{matrix}	Error	$V_{lesion\ 1}$	$CV_{lesion\ 1}$	Error	$V_{lesion\ 2}$	$CV_{lesion\ 2}$	Error	$V_{lesion\ 3}$	$CV_{lesion\ 3}$	Error
1	194.9	1.02	1.82	0.28	3.45	10.82	0.19	5.11	11.42	0.07	6.90	27.84
2	193.1	2.01	3.01	0.28	6.15	11.52	0.18	5.25	13.68	0.08	7.70	25.50
3	197.6	1.73	1.51	0.31	5.68	5.10	0.20	2.53	6.30	0.07	6.90	27.84
4	195.8	1.18	1.51	0.30	6.09	6.04	0.20	7.07	7.53	0.09	15.19	18.71
Mean	195.3	1.49	1.96	0.29	5.34	8.37	0.19	4.99	9.73	0.08	9.17	24.97

ducibility of the technique must be firmly established. Studies of feature map segmentation accuracy were reported by Cline et al. (18) using a multicompartment gel phantom designed to mimic the transverse section of the human head and Kohn et al. (17) using relatively large graphite-containing gel pieces in a gel matrix. Both of these studies used phantoms exhibiting high contrast changes as compared with the present study. Using nearest neighbor cluster analysis for feature map generation, Cline et al. (18) compared computed areas with known areas in a single slice and reported percent errors in areas ranging from −6 to +7%, which corresponded to −9 cm^2 for the larger "brain" compartment and +0.2 cm^2 for a smaller "tumor" compartment. Kohn et al. (17), using semiautomated line partitioning of bifeature space, obtained an RMS percent error of ~9% for volumes ranging from 4.5 to 32.5 cm^3, and the overall RMS error in volume was determined to be 0.91 cm^3. From the results of the present study, the feature map segmentation technique yielded volumes accurate to 0.03 cm^3 for the smaller lesion gels and 3.8 cm^3 for the larger matrix volume. Together, these results indicate the potential for measurements of normal and pathological structures in human brain that are accurate to within 1 cm^3. Furthermore, the exceedingly small volume of the 0.10 cm^3 lesion gel establishes the fact that this technique can accurately segment such small volumes even with the rather minimal contrast between the lesion and matrix gels (Fig. 1a and b).

In terms of reproducibility, our phantom data indicate that, on the average, a given observer can be expected to report volumes with a coefficient of variation of 6.5% for small volumes such as the lesion gels, but this variation decreases to 1.5% for the larger matrix gel volume. Furthermore, different observers can be expected to report volumes that vary by 5.7% for small volumes and by 1.0%

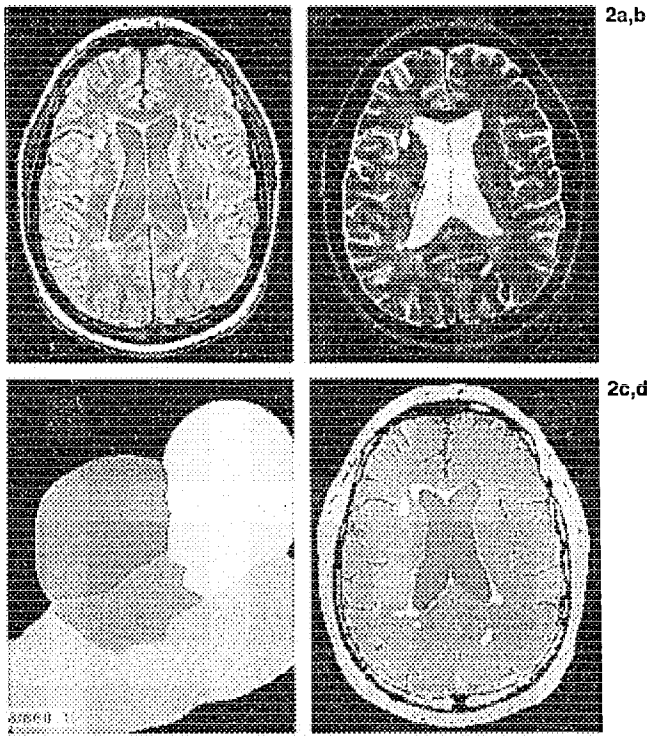

FIG. 2. One image plane taken from the image data of an MS patient. **a:** Proton density-weighted image. **b:** T2-weighted image. **c:** Feature map. **d:** Segmented image. For the feature map and segmented image, blue represents CSF, gray represents brain, yellow represents MS lesions, and pink represents other tissue (scalp, marrow, etc.).

TABLE 2. *Computed volumes (V; cm³) from MS patient data*

Patient no.	V_{brain} ± SD	V_{CSF} ± SD	$V_{lesions}$ ± SD
1	1,240 ± 90	114 ± 24	22 ± 4
2	1,298 ± 39	159 ± 32	35 ± 8
3	1,279 ± 50	156 ± 21	26 ± 2
4	1,060 ± 77	203 ± 25	73 ± 15
5	1,297 ± 61	114 ± 19	24 ± 6

Values shown are averaged over four repetitions from four observers.

for larger volumes. It should be mentioned that these values were obtained from observers with a wide range of experience with respect to the segmentation algorithm and the variation decreases significantly with increased experience level.

Two-dimensional feature map segmentation for volumetric analysis has been applied to MS patient image sets previously by Herman (20). In that study of three MS patients, image sets were segmented two times by each of three observers with variable experience. Average brain and MS lesion intraobserver variabilities were 1.7 and 14.5%, respectively, and the interobserver variabilities were 1.6 and 16.4%, respectively. To compare these values to the coefficients of variation reported in the present study, the variabilities were converted to coefficients of variation. The intraobserver coefficients of variation obtained were 2.4 and 20.6% for brain and lesion, respectively, and the interobserver coefficients of variation were 2.6 and 33.9% for brain and lesion, respectively. No quantitation of CSF was reported. Comparison of these data with those obtained in the present study indicates that although the reproducibilities of brain volumes are generally comparable, the interobserver coefficient of variation for MS lesions was significantly less in the present studies. This might well be explained by the thinner image slices acquired in the present studies (3 vs. 5 mm), yielding decreased partial volume averaging and improved lesion definition.

It should be noted that preliminary studies indicate that reproducibility might be significantly enhanced by preprocessing the image data with an anisotropic diffusion filter (26) that averages (decreases noise) in low frequency regions of the image without degrading high frequency edges. As recently reported by Gerig et al. (27), the application of such a filter improves the cluster definitions in feature space and results in more accurate tissue segmentation. Additionally, although the percent of misclassified pixels was low, most of these pixels were in regions such as scalp, bone marrow, and facial tissue. Elimination of such extrameningeal tissues by means of three-dimensional connectivity algorithms further improves the accuracy and reproducibility of the segmentation. Application of the diffusion filter and connectivity algorithm to three of the five patient data sets listed in Table 2 decreased the interobserver variation in MS lesion volume determinations by a factor of 2 (four observers, four repetitions). The CSF and brain volume reproducibilities were also improved by factors of 2 and 3, respectively. Based on these preliminary results, nonparametric feature map segmentation of such preprocessed image sets should provide for volumetric determinations of brain, CSF, and MS lesions that are reproducible to within 1, 7, and 10%, respectively, for both inter- and intraobserver. Furthermore, such preprocessing allows for improved segmentation of gray matter and normal-appearing white matter in addition to the lesions and CSF (unpublished data). A more detailed study of preprocessing techniques is currently underway in patients and normal volunteers.

Finally, although the present results consider total tissue volumes (summed over all sections), it is straightforward to obtain per-section lesion volumes or, with postsegmentation connectivity processing, to obtain individual lesion volumes. Previously, per-lesion volume determinations have been performed using manual tracing techniques as discussed by Goodkin et al. (8). As previously mentioned, however, these techniques generally are quite time consuming and are usually performed by a single operator to prevent large interobserver variations. Therefore, such techniques are difficult to use routinely for monitoring large patient populations.

CONCLUSIONS

As noted above, the feature map segmentation technique offers several advantages for semiautomated volumetric analysis of MS patient image data. First, operator interaction is limited to sampling of the tissue types of interest. Therefore, operator bias is decreased. Second, segmentation is performed relatively quickly, requiring only 20–30 min for the complete segmentation of ≈50 dual-echo image planes using a Sun Microsystems SPARCstation2 workstation. Third, as these studies have shown, the segmentation can be performed by individuals without extensive training and still yield acceptable reproducibility and accuracy. Finally, the nature of the feature map constructed by this nonparametric analysis can account for image shading due to RF inhomogeneities as long as the full range of intensities representing each tissue is sampled (17). The accuracy and reproducibility of the technique, particularly when coupled with appropriate preprocessing of the image data as described above, promise to provide a highly quantitative index of MS lesion burden, as well as brain and CSF volumes, which should greatly aid in the

initial evaluation and follow-up of MS patients. Furthermore, the segmented data can be displayed "three-dimensionally" using surface or volume rendering techniques (28), providing greatly improved visualization of lesion distribution.

Acknowledgment: This investigation was supported in part by a grant from the National Multiple Sclerosis Society and the John S. Dunn Foundation.

REFERENCES

1. Paty DW, Bergstrom M, Palmer M, MacFadyen J, Li D. A quantitative magnetic resonance image of the multiple sclerosis brain. *Neurology* 1985;35(suppl 1):137.
2. Paty DW. Magnetic resonance imaging in the assessment of disease activity in multiple sclerosis. *Can J Neurol Sci* 1988; 15:266–72.
3. Cramer GD, Allen DJ, DiDio LJA, Potvin W, Brinker R. Evaluation of encephalic ventricular volume from the magnetic resonance imaging scans of thirty-eight human subjects. *Surg Radiol Anat* 1990;12:287–90.
4. Young AH, Blackwood DHR, Roxborough H, McQueen JK, Martin MJ, Kean D. A magnetic resonance imaging study of schizophrenia: brain structure and clinical symptoms. *Br J Psychiatry* 1991;158:158–64.
5. Kesslak JP, Nalcioglu O, Cotman CW. Quantification of magnetic resonance scans for hippocampal and parahippocampal atrophy in Alzheimer's disease. *Neurology* 1991;41: 51–4.
6. Hynd GW, Semrud-Clikeman M, Lorys AR, Novey ES, Eliopulos D, Lyytinin H. Corpus callosum morphology in attention deficit-hyperactivity disorder: morphometric analysis of MRI. *J Learn Disabil* 1991;24:141–6.
7. Liu CK, Miller BL, Cummings JL, et al. A quantitative MRI study of vascular dementia. *Neurology* 1992;42:138–43.
8. Goodkin DE, Ross JS, Vanderbrug-Medendorp S, Konecsni J, Rudick RA. Magnetic resonance imaging lesion enlargement in multiple sclerosis. *Arch Neurol* 1992;49:261–3.
9. Pfefferbaum A, Lim KO, Rosenbloom M, Zipursky RB. Brain magnetic resonance imaging: approaches for investigating schizophrenia. *Schizoph Bull* 1990;16:453–76.
10. Rusinek H, de Leon MJ, George AE, et al. Alzheimer disease: measuring loss of cerebral gray matter with MR imaging. *Radiology* 1991;178:109–14.
11. Jernigan TL, Archibald SL, Berhow MT, Sowell ER, Foster DS, Hesselink JR. Cerebral structure on MRI, part I: localization of age-related changes. *Biol Psychiatry* 1991;29:55–67.
12. DeCarli C, Maisog J, Murphy DGM, Teichberg D, Rapoport SI, Horwitz B. Method for quantification of brain, ventricular, and subarachnoid CSF volumes from MR images. *J Comput Assist Tomogr* 1992;16:274–84.
13. Jack CR Jr, Petersen RC, O'Brien PC, Tangalos EG. MR-based hippocampal volumetry in the diagnosis of Alzheimer's disease. *Neurology* 1992;42:183–8.
14. Pannizzo F, Stallmeyer MJB, Friedman J, et al. Quantitative MRI studies for assessment of multiple sclerosis. *Magn Res Med* 1992;24:90–9.
15. Vannier MW, Butterfield RL, Jordan D, Murphy WA, Levitt RG, Gado M. Multispectral analysis of magnetic resonance images. *Radiology* 1985;154:221–4.
16. Cline HE, Lorensen WE, Kikinis R, Jolesz F. Three-dimensional segmentation of MR images of the head using probability and connectivity. *J Comput Assist Tomogr* 1990; 14:1037–45.
17. Kohn MI, Tanna NK, Herman GT, et al. Analysis of brain and cerebrospinal fluid volumes with MR imaging. Part I. Methods, reliability, and validation. *Radiology* 1991;178: 115–22.
18. Cline HE, Lorensen WE, Souza SP, et al. 3D surface rendered MR images of the brain and its vasculature. *J Comput Assist Tomogr* 1991;15:344–51.
19. Tanna NK, Kohn MI, Horwich DN, et al. Analysis of brain and cerebrospinal fluid volumes with MR imaging: impact on PET data correlation for atrophy. Part II: aging and Alzheimer dementia. *Radiology* 1991;178:123–30.
20. Herman GT. Quantitation using 3D images. In: Udupa JK, Herman GT, eds. *3D imaging in medicine*. Boca Raton: CRC Press, 1991:145–61.
21. Rao SM, Leo GJ, Haughton VM, St. Aubin-Faubert P, Bernadin L. Correlation of magnetic resonance imaging with neuropsychological testing in multiple sclerosis. *Neurology* 1989;39:161–6.
22. Hart PE. The condensed nearest neighbor rule. *IEEE Trans Inform Theory* 1968;IT14:515–6.
23. Duda RO, Hart PE. *Pattern classification and scene analysis*. New York: Wiley, 1973:85–126.
24. Companion JA. Tissue-simulating gel for medical research. *NASA Tech Briefs* 1992;80.
25. Cline HE, Dumoulin CL, Hart HR Jr, Lorensen WR, Ludke S. 3D reconstruction of the brain from magnetic resonance images using a connectivity algorithm. *Magn Res Imag* 1987;5:345–52.
26. Perona P, Malik J. Scale-space and edge detection using anisotropic diffusion. *IEEE Trans Patt Anal Mach Intell* 1990;12:629–39.
27. Gerig G, Kübler O, Kikinis R, Jolesz FA. Nonlinear anisotropic filtering of MRI data. *IEEE Trans Med Imag* 1992;11: 221–32.
28. Udupa JK. Computer aspects of 3D imaging in medicine: a tutorial. In: Udupa JK, Herman GT, eds. *3D imaging in medicine*. Boca Raton: CRC Press, 1991:1–69.